Supporting Information

Edayathumangalam et al. 10.1073/pnas.1315887110

SI Materials and Methods

Protein Expression and Purification. The full-length coding region of the *Bacillus subtilis gabR* gene (residues 1–479) was subcloned from a pBAD construct into pETite C-His vector (Lucigen), which encodes an inducible T7 expression system. The C-terminally His_6-tagged protein was expressed in *Escherichia coli* strain BL21 (DE3). Transformed cells were grown in LB media containing kanamycin at 37 °C to OD_{600} of ∼0.6 and induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37 °C for 4 h. Cells were harvested, resuspended in buffer containing twofold PBS buffer and 10 mM imidazole (pH 7.5), and sonicated. Insoluble material was removed by centrifugation. The GabR-His_6 protein (molecular weight, 56 kDa) was purified to homogeneity by affinity chromatography Ni-NTA (GE Healthcare) and gel filtration (HiLoad 16/60 Superdex 200 pg; GE Healthcare) in buffer containing twofold PBS (274 mM NaCl, 5.4 mM KCl, 16 mM Na_2HPO_4, 2.92 mM KH_2PO_4, pH 7.4), 475 mM imidazole (pH 7.5), 1 mM pyridoxal 5′-phosphate (PLP), and 5% (vol/vol) glycerol. The protein requires high salt and imidazole concentrations to stay soluble without precipitation. Alternatively, the presence of 5% (vol/vol) glycerol and 0.5% Nonidet P-40 in the buffer helps to keep GabR soluble without the need for NaCl and imidazole. Purified GabR-His_6 protein was concentrated (Centricon Plus centrifugal filter units; Millipore) and stored at 4 °C for short term (∼4 wk) or flash frozen in ∼50-μL aliquots in liquid nitrogen after the addition of 15% (vol/vol) glycerol and stored at −80 °C for long-term use. The concentrated sample of purified GabR was pale yellow in color, indicating that PLP was bound to the protein.

Analytical Ultracentrifugation Analysis. The oligomeric state of GabR−PLP in solution was analyzed by sedimentation velocity using a Beckman-Coulter Optima XL-I ultracentrifuge. The experiments were conducted at low protein concentrations (30, 50, and 80 μg/mL) in the presence of 0.5 mM PLP and in the presence and absence of 20 mM GABA. In each experiment, ∼450 μL of the protein solution was loaded into a sector-shaped centrifugation cell at concentrations corresponding to OD_{230} values of 0.3, 0.5, and 0.8. The samples were spun at 48,000 rpm (Beckman AN60Ti Rotor) for 5 h with radial scans capturing the boundary distribution every 2 min until the sample had almost completely pelleted. Boundary profiles from sedimentation velocity scans were collected and analyzed using the UltraScan II software suite. Initial model-independent van Holde–Weischet analysis of the data clearly showed no evidence of mass action over the concentration range investigated, thereby confirming an absence of dynamic equilibrium between species. Therefore, all data were fit to a 50-iteration global genetic algorithm–Monte Carlo (GA-MC50) model, which showed a narrow and randomly distributed residual, indicative of good data quality and model fitting.

Fluorescence Polarization DNA-Binding Assay. Fluorescence polarization data for the GabR–DNA complex were obtained using an EnVison 2102 Multilabel Plate Reader (Perkin-Elmer). Two 49-bp complementary single-stranded DNA oligonucleotides containing the 47-bp GabR binding region (5′-TCTGATACCATCAAAAAGTTATAATTGGTACTTTTCATCATACCAAAGA-3′) were purchased from Integrated DNA Technologies. The oligonucleotides were either unlabeled or labeled at the 3′ end with the fluorophore, TAMRA. The double-stranded oligonucleotides were prepared by annealing a mixture containing equimolar concentrations of the two single-stranded oligonucleotides. For the fluorescence polarization assay, varying concentrations of the GabR protein was titrated against a fixed concentration of TAMRA-labeled DNA (50 nM). GabR protein was serially diluted in the DNA binding buffer [20 mM Tris·HCl, pH 8.0, 50 mM KCl, 2 mM $MgCl_2$, 5% (vol/vol) glycerol, 1 mM EDTA, 1 mM DTT, and 0.05% Nonidet P-40]. The TAMRA-DNA (50 nM) was mixed with the serially diluted GabR protein (final concentration, 0–1,000 nM) and 100 μL of each sample mixture was loaded into the individual wells of a 384-well microplate. In the competitive binding assay, varying concentrations of unlabeled DNA were titrated against a fixed concentration of preformed GabR-labeled DNA complex (100 nM GabR and 100 nM TAMRA-DNA).

The fluorescence polarization (FP) signals of the GabR−DNA mixtures were measured with fixed excitation (531 nm) and emission (595 nm) wavelength filters at room temperature. Fluorescence intensity remained constant throughout the polarization measurements. The fluorescence polarization signal for each sample was calculated and plotted against the GabR protein concentrations in the fluorescence polarization assay and against the concentrations of unlabeled DNA in the competitive binding assay. Each data point is the average of three independent measurements. The titration plot for GabR versus TAMRA-DNA was fitted to a modified Hill equation (Eq. **S1**) in ORIGIN software to calculate the dissociation constants and Hill coefficient. For the competitive binding assay, the IE_{50} value and Hill coefficient were obtained by fitting the data points to the Hill equation (Eq. **S2**).

$$(FP_{GTD} - FP_{FTD}) = [\text{TAMRA-DNA}]^n \Big/ \left(K_D^n + [\text{TAMRA-DNA}]^n\right), \quad \textbf{[S1]}$$

where FP_{GTD} represents FP signal when TAMRA-DNA is saturated with GabR; FP_{FTD} represents FP when all TAMRA-DNA is free in solution without any GabR bound; K_D is the dissociation constant for the GabR–TAMRA-DNA complex; and n is the Hill coefficient.

$$(FP_{GTD} - FP_{FTD}) = [\text{Unlabeled DNA}]^n \Big/ \left(IE_{50}^n + [\text{Unlabeled DNA}]^n\right), \quad \textbf{[S2]}$$

where FP_{GTD}, FP_{FTD}, and n are the same as in Eq. **S1**; IE_{50} represents the concentration of unlabeled DNA needed to bind 50% of protein.

Crystallization. Crystals of GabR were obtained by the hanging-drop method using protein at 5 or 10 mg/mL The initial sparse matrix screen was set up with the Gryphon crystallization robot (Art Robbins Instruments) to obtain initial crystallization hits, which were further optimized manually to obtain crystals with the best size and morphology.

The optimized crystallization solution contained 33% (vol/vol) PEG400, 200 mM calcium acetate, and 200 mM imidazole, pH 7.5. The hanging drops were incubated at 20 °C or room temperature. Crystals appeared in 5 d and grew to maximum size in 1–2 wk. Crystals with good morphology and size were picked directly from the hanging drops and flash-cooled in liquid nitrogen. The heavy metal derivatives were prepared by soaking crystals in 10 mM ethylmercury phosphate for 1 h before flash-

cooling in liquid nitrogen. The heavy metal soaking condition also included 1 mM PLP.

Data Collection and Processing. Single-wavelength anomalous dispersion (SAD) data were collected at beamlines 19-ID and 19-BM at Advanced Photon Source, Argonne National Laboratory. X-ray diffraction data for the mercury derivative of holo-GabR (Table S1, Dataset 1) and for holo- and apo-GabR native crystals (Table S1, Datasets 2 and 3) were collected at the wavelengths of 1.0039 and 0.9792 Å, respectively. The datasets were indexed, integrated, and scaled with either the HKL package (1) or iMosflm and Scala programs in the CCP4 program suite (2). The datasets were all scaled in the space group $P2_12_12_1$ to resolutions of 3.1, 2.7, and 2.55 Å, respectively. The data statistics are reported in Table S1.

Sequence and Structure Analyses. GabR homologs were identified in the sequence database using BLAST (3); multiple sequence alignment using BLAST; conserved domain searches using CDD (4); buried surface area from PDBePISA server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver); and structural similarity using DALI (5). Secondary structure superimpositions were done in Coot. Electrostatic surface potentials were calculated with PyMOL (The PyMOL Molecular Graphics System, version 1.5.0.4; Schrödinger). Figs. 3, 4, 6, and 9, and Figs. S2, S4, S5, S7, and S8 were drawn with PyMOL.

Structure Solution, Model Building, and Refinement. Automated substructure solution by AutoSol in PHENIX (6) found 15 heavy atom peaks in the mercury dataset (Table S1, Dataset 1) and identified fourfold noncrystallographic symmetry (NCS) using superposition of the mercury sites. AutoBuild in PHENIX automatically built ∼80% of the residues. The initial phases were extended and refined to 2.7 Å using the native holo-GabR dataset (Table S1, Dataset 2) to yield improved maps and further improved by fourfold NCS averaging and density modification. The remainder of the model was built manually in Coot (7) guided by the structure of *Thermococcus profundus* multiple substrate aminotransferase [Protein Data Bank (PDB) ID code 1WST], a homolog obtained from the FFAS03 server (http://ffas.burnham.org/ffas-cgi/cgi/ffas.pl). Several loops accounting for gaps in the model were placed and refined by a combination of manual building and the *fit_loops* routine in PHENIX. Next, MR-SAD using the model and the mercury dataset identified five more mercury sites and led to even better maps. Several iterative rounds of manual model building and refinement using the maximum-likelihood amplitude targets in PHENIX and Rosetta using simulated annealing (in the initial stages) and automatic NCS restraints were required to obtain the final model. The initial model (∼80% complete) had R_{cryst} and R_{free} values of 0.393 and 0.484, respectively, for all data between 45.0 and 2.7 Å. The R_{cryst} and R_{free} values of the final model (98% complete) are 0.229 and 0.265, respectively. The final model of the GabR−PLP complex consists of two homodimers of GabR in the asymmetric unit with ∼470 residues out of 479 accounted for in each monomer. The final model also contains 1 molecule of PLP per monomer, 16 molecules of ethylmercury, and 52 water molecules. The quality of the final model is summarized in Table S1. This model has been deposited in the PDB and is available with ID code 4N0B.

The native dataset of apo-GabR (Table S1, Dataset 3), which was processed to a resolution of 2.55 Å, was phased by molecular replacement using the final model of GabR as the input model. The R_{cryst} and R_{free} values of the final model from the native dataset are 0.195 and 0.246, respectively. Other model statistics are summarized in Table S1. This model has been deposited in the PDB with ID code 4MGR. The density for PLP is lacking in this structure, as expected. We interpreted a difference density peak in the PLP-binding pocket as imidazole (Fig. S3) because imidazole is a major component of the protein purification and crystallization buffers.

GabR Enzyme Assay. Activity was measured using varying concentrations of GabR protein in an assay buffer containing 100 mM Tris·HCl, 5 mM DTT, 100 mM KCl, 100 μM PLP, 20 mM GABA, 20 mM α-ketoglutarate, 500 μM $NADP^+$, and 3 units of aldehyde dehydrogenase (ADH). In the coupled assay, ADH consumes the product of the first half-reaction (succinic semialdehyde) to produce succinate. The coupled reaction is measured spectrophotometrically (absorbance change at 340 nm) from the concomitant reduction of the cofactor, $NADP^+$ to NADPH.

1. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, eds Carter CW, Sweet RM (Academic, New York), Vol 276, pp 307–326.
2. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5): 760–763.
3. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
4. Marchler-Bauer A, et al. (2011) CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* 39(Database issue):D225–D229.
5. Holm L, Rosenstrom P (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res* 38(Web Server issue):W545–W549.
6. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
7. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Sp Iss 1):2126–2132.

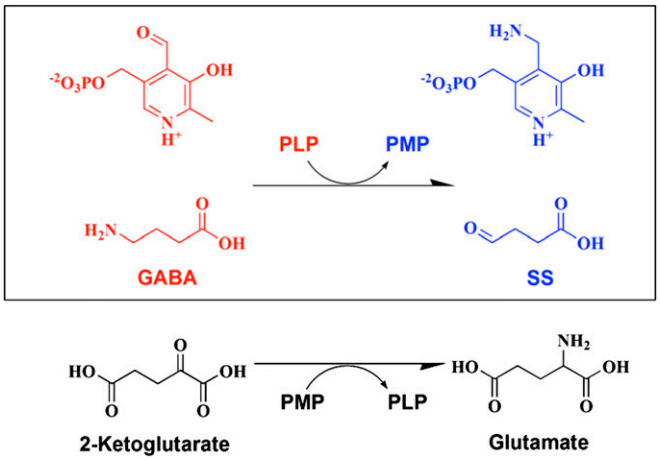

Fig. S1. The GabT-catalyzed aminotransferase reaction. The first half-reaction (GABA to succinic semialdehyde) of the GabT-catalyzed ping-pong transamination is shown in the boxed area. The second half-reaction is shown outside the boxed area.

Fig. S2. Quaternary structure of GabR. The structure is rotated 90° counterclockwise along the *y* axis relative to Fig. 2.

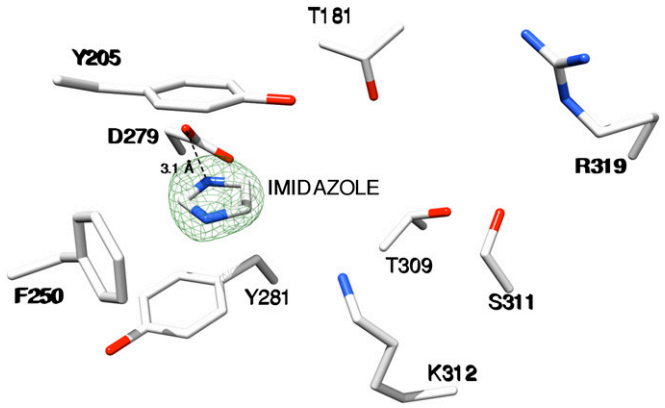

Fig. S3. Imidazole is bound in the PLP-binding pocket of apo-GabR. The PLP-binding site of the crystal structure of apo-GabR is shown. The $2F_o - F_c$ map (green mesh) around the imidazole molecule is shown at 1σ level. The imidazole moiety and surrounding side chains are shown as sticks. Carbon atoms are shown in white, oxygen atoms in red, and nitrogen atoms in blue.

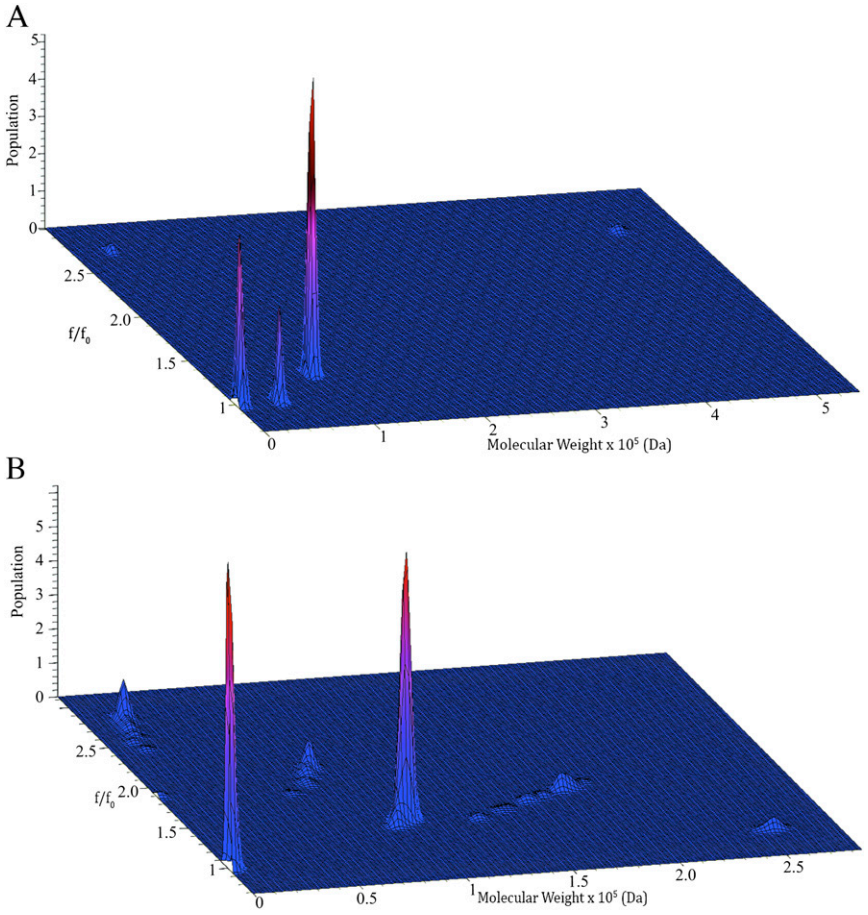

Fig. S4. Sedimentation velocity analysis of GabR. (*A*) Molecular weight of the GabR–PLP sample determined from the global fitting of sedimentation velocity data. (*B*) Molecular weight of the GabR–PLP sample in the presence of 20 mM GABA.

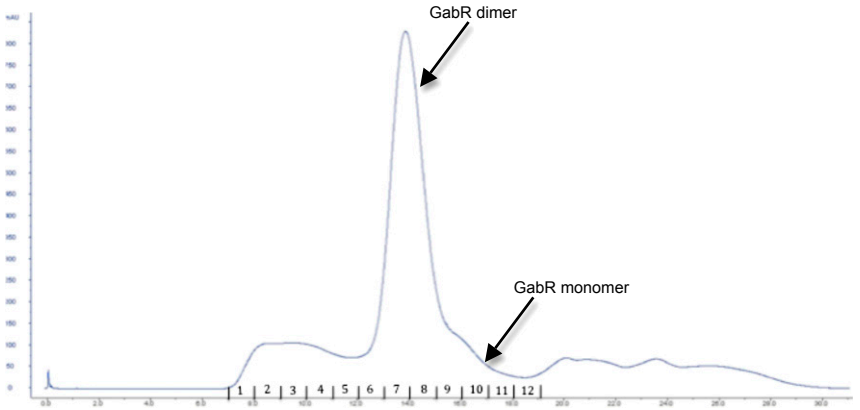

Fig. S5. The gel-filtration profile of GabR. The major peak for GabR (fractions 7–9) on Superdex 200 16/60 corresponds to an apparent molecular weight of a dimer. The minor peak (fractions 9–10) corresponds to that for a monomer.

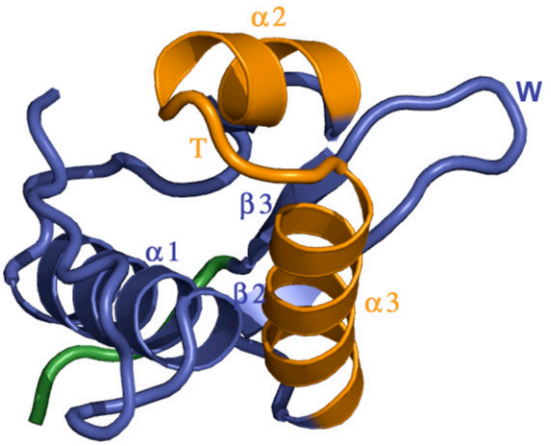

Fig. S6. The GabR N-terminal winged-helix domain. The GabR N-terminal winged-helix domain (residues 3–80) is shown in detail. The various secondary structural elements (α1-β1-α2-T-α3-β2-W-β3) are labeled. The helix-turn-helix (HTH) motif (α2-T-α3) is shown in orange. W is the wing (β2-W-β3).

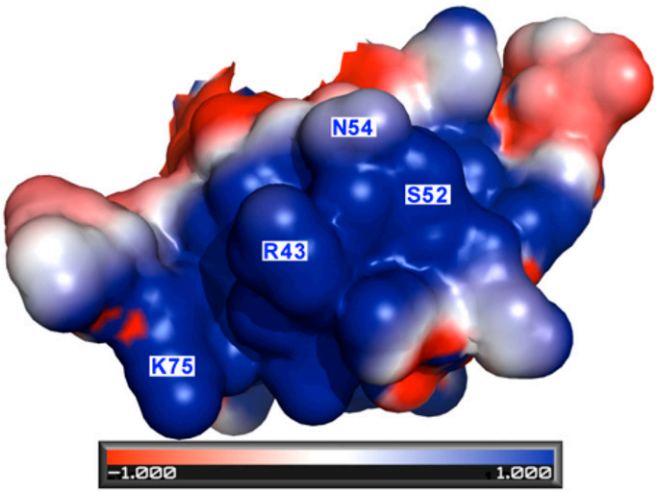

Fig. S7. The most basic surface on the GabR winged-helix domain. The electrostatic potential surface of the GabR winged-helix domain is shown. The values −1.000 and 1.000 on the scale bar represent acidic and basic electrostatic potential surface charges, respectively. The orientation corresponds to the winged-helix domain shown on the bottom in Fig. 2. The labeled residues denote amino acids that are conserved between GabR and FadR.

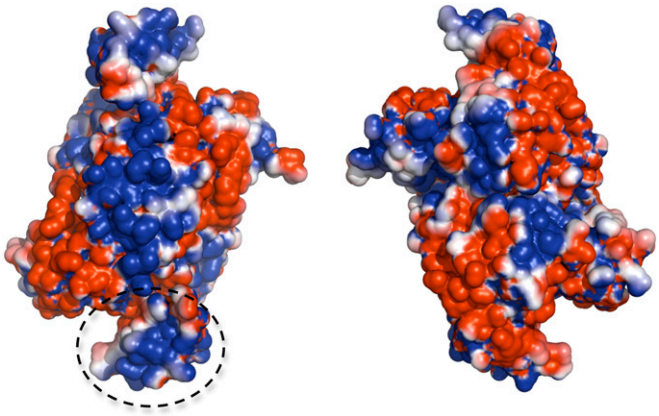

Fig. S8. GabR electrostatic surface potential. The electrostatic surface potential representation for the GabR homodimer in two different views. The figure on the *Left* is in the same orientation as Fig. 2. The dotted circle highlights the most basic surface on the winged-helix domain. The figure on the *Right* is rotated 90° counterclockwise relative to the figure on the *Left*.

Fig. S9. Sequence alignment of the winged-helix domain sequences of *B. subtilis* GabR and *E. coli* FadR. A BLAST alignment of the HTH and wing motifs are shown. The green asterisks highlight DNA-binding residues of FadR that are identical or similar in GabR.

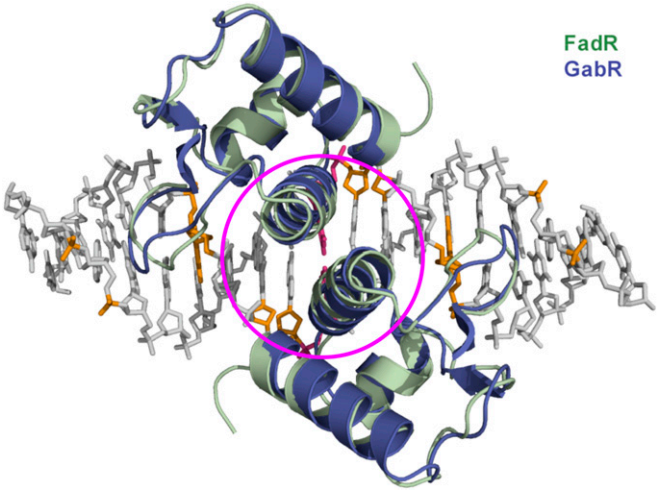

Fig. S10. Two DNA-binding domains of GabR can interact on DNA through the "α3–α3 dimer interface." A superposition of two molecules of GabR onto the FadR–DNA complex is shown. The α3–α3 interface is indicated by the magenta circle. The acyl-CoA–binding domains of FadR and putative aminotransferase domains of GabR are not shown for clarity. The protein and DNA are shown in cartoon and stick representation, respectively. Side chains of some of the DNA-binding residues are shown as sticks and labeled. The corresponding phosphate groups and nucleotides contacted by the protein side chains are colored in orange. The central DNA base pair is shown in pink. Only the winged-helix domains of FadR are shown. The acyl-CoA–binding domains of FadR are hidden for clarity.

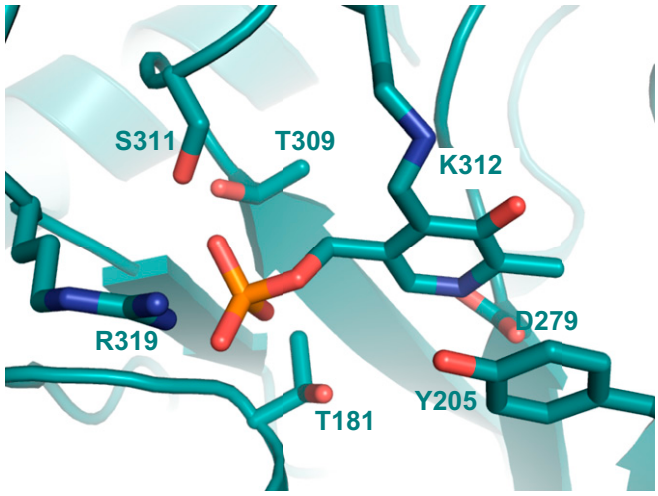

Fig. S11. The PLP-binding pocket of *B. subtilis* GabR. The GabR residues that interact with PLP are labeled.

```
35   KGEGAELYDLDGRRFIDFAGAIGTLNVGHSHPKVVEAVKRQ---AEELIHPGFNVMMYPT  91
     +  G  L D+DG R +D    I ++ +G+SHP +V+ V++    +  +  P   ++
51   ESRGNYLVDVDGNRMLDLYSQISSIPIGYSHPALVKLVQQPQNVSTFINRPALGILPPEN  110

92   YIE-LAEKLCGIAPGSHEKKAIFLNSGAEAVENAVKIARKY--TKRQG-----------  136
     ++E L E L  +AP    +  I +  G+ + ENA K    +  +K +G
111  FVEKLRESLLSVAPKGMSQ-LITMACGSCSNENAFKTIFMWYRSKERGQSAFSKEELETC  169

137  ------------VVSFTRGFHGRTNMTMSMTSKVKPYKFGFGPFAPEVYQAPFPY--YYQ  182
                 ++SF   FHGRT   ++ T     +K     F   +  APFP   Y
170  MINQAPGCPDYSILSFMGAFHGRTMGCLATTHSKAIHKIDIPSFDWPI--APFPRLKYPL  227

183  KPAGMSDESYDDMVIQAFNDFFIA-SVAPETVACVVMEPVQGEGGFIIPSKRFVQHVASF  241
     +     ++  +   ++   D  +      +TVA +++EP+Q EGG    S  F + +
228  EEFVKENQQEEARCLEEVEDLIVKYRKKKKTVAGIIVEPIQSEGGDNHASDDFFRKLRDI  287

242  CKEHGIVFVADEIQTGFARTGTYFAIEHF--DVVPDLITVSKSLAAGLPLSGVIGRAEML  299
      ++HG  F+ DE+QTG   TG ++A EH+  D   D++T SK +     ++G     E
288  SRKHGCAFLVDEVQTGGGSTGKFWAHEHWGLDDPADVMTFSKKM-----MTGGFFHKEEF  342

300  DAAAPGELGGTYAGSPLGCAAALAVLDIIEEEGLNERSEEIGKIIEDKAYEWKQEFP-FI  358
        AP  +  T+ G P        V++II+ E L   +   GK++     + +  +P FI
343  RPNAPYRIFNTWLGDPSKNLLLAEVINIIKREDLLSNAAHAGKVLLTGLLDLQARYPQFI  402

359  GDIRRLGAMAAIEIVKDPDTREPDKTKAAAIAAYANQNGLLLLTAGINGNIIRFLTPLVI  418
       +R  G   + +         PD++    + + A   G++L   G     IRF   LV
403  SRVRGRGTFCSFDT--------PDESIRNKLISIARNKGVML--GGCGDKSIRFRPTLVF  452

419  SDSLLNEGLSIL  430
      D   +  L+I
453  RDHHAHLFLNIF  464
```

Fig. S12. BLAST protein sequence alignment. Sequence alignment between *B. subtilis* GabT GABA aminotransferase (upper sequence; 1–436) and pig liver GABA-AT (lower sequence; 1–472) are shown. The two proteins share 25% sequence identity and 43% similarity between their AT domains. The conserved residues interacting with the substrate carboxylate groups (R192 and R445), the pyridinium N and phenolic O of PLP (D298 and Q301, respectively), and the invariant Schiff base lysine (K329) are highlighted. The residue numbers in parentheses denote residues from pig liver GABA-AT.

Table S1. Data collection and refinement statistics

Statistics	Datasets		
	(1) GabR-PLP, mercury derivative	(2) GabR-PLP, native	(3) apo-GabR, native
Wavelength, Å	1.0039	0.9792	0.9792
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Resolution range, Å	30–3.1 (3.15–3.1)	50–2.71 (2.76–2.71)	49.3–2.55
Unit cell (*a b c*), Å	98.1 101.1 211.2	97.2 101.3 211.2	99.9 101.4 213.2
Total unique reflections	39,106	57,413	68,270
Multiplicity	6.7 (6.8)	4.8 (3.8)	5.8 (5.6)
Completeness, %	100 (100)	99 (97.9)	95.8 (90.7)
Linear R_{merge}*, %	10.7 (71.6)	0.102 (ND)	11.7 (68.7)
I/σ	21.9 (2.9)	18.7 (1.0)	9.4 (2.2)
Average Wilson B factor	67.7	59.9	38.7
R_{cryst}[†]	0.2146		0.1942
R_{free}[‡]	0.2565		0.2437
Number of non-H atoms	15,233		15,770
Macromolecules	15,167		15,230
Ligands	11		73
Water	55		467
RMSD[§] bonds, Å	0.002		0.004
RMSD[§] angles, °	0.55		1.125
Ramachandran favored, %	93.8		93.0
Ramachandran outliers, %	1.1		0.3
Clashscore	4.68		10.80
Average B factor, Å^2	77.8		46.8
Macromolecules	77.9		47.2
Solvent	49.8		36.0

The values for the highest resolution bin are in parentheses. ND, not determined/calculated by the program.

*Linear $R_{merge} = \sum |I_{obs} - I_{avg}| / \sum I_{avg}$.

[†]$R_{cryst} = \sum |F_{obs} - F_{calc}| / \sum F_{obs}$.

[‡]Five percent of the reflection data were selected at random as a test set and only these data were used to calculate R_{free}.

[§]RMSD, root-mean-square deviation.